A PSYCHOACOUSTIC AUDITORY DISPLAY FOR NAVIGATION

Tim Ziemer

University of Bremen
Bremen Spatial Cognition Center
Enrique-Schmidt-Str. 5, 28359 Bremen, Germany
ziemer@uni-bremen.de

Holger Schultheis

University of Bremen
Bremen Spatial Cognition Center
Enrique-Schmidt-Str. 5, 28359 Bremen, Germany
schulth@uni-bremen.de

ABSTRACT

A psychoacoustic auditory display for navigation in two-dimensional space is presented. The auditory display is examined in an experiment with novice users. Trajectory analysis indicates that users were able to a) accurately find sonified targets b) analyze the sonification axis-by-axis c) integrate the sonified dimensions to approach the target on the shortest path. Techniques developed in this work appear to work equally well with three-dimensional coordinates.

1. INTRODUCTION

Many specialized operations have in common that human-machine interaction is supported by visual navigation assistance. Examples include car parking, piloting, remote vehicle control and minimally invasive surgery. Auditory displays have been proposed as complement or even as alternative for visual assistance in such operations [1, 2, 3, 4]. However, these auditory displays communicate only sparse information about the spatial context.

In the present study, a novel, psychoacoustic auditory display is derived, implemented, and examined by means of an experiment with 18 novice users. It acts as a standalone-solution for navigation in 2-dimensional space, without the need for additional visualization.

2. BACKGROUND AND RELATED WORK

A major problem associated with multi-dimensional auditory displays is ambiguity due to perceptual interactions between orthogonal physical audio parameters [5]. For example, physical frequency affects both perceived loudness and pitch and even physical amplitude may affect both loudness and pitch perception [6]. Consequently, these physically independent parameters cannot serve to communicate orthogonal dimensions to a human operator, because they are not orthogonal in perception. It has been observed that a persisting problem in auditory display research is that auditory display design is often arbitrary or based on engineering convenience, and the need to thoroughly consider auditory perception, e.g. in terms of psychoacoustics, has been expressed [7, 8, 9, 10, 11, 12, 13, 14, 15, 16]. An overview about psychoacoustics is given in [6].

Some theoretic approaches towards psychoacoustic sonification can be found in [8, 17, 12]. The first leverages pitch as height, timbre as angle, and brightness as radius in a cylindrical coordinate system, following the example of color space. [8] points out the distinction between objective physical and subjective perceptual aspects of sound and identifies several obstacles for the conduct of his approach. [17] define a framework that translates physical gesture input over gesture-perception and auditory perception to physical audio output. They identify potential psychoacoustical parameters, such as pitch, loudness, timbre aspects, like brightness, roughness, vibrato and formants, as well as their temporal evolution in terms of derivatives. [12] identify loudness, sharpness, roughness, beating and pitch as potential psychoacoustic parameters for perceptual sonification and suggest to map two input data streams to two sensations for two-dimensional sonification. However, the authors realized that finding the right audio parameters that create the desired perceptual outcome is an inverse problem. They suggest lookup tables to find one possible constellation of audio parameter settings to create the desired perceptual sound impression. The risk here is that continuous changes in the data input may cause sudden jumps of audio parameters.

From these studies we derive three critical demands on auditory displays for navigation:

- (α) The auditory display is interactively interpretable
- (β) The auditory parameters of each axis are orthogonal in perception
- (γ) The auditory axes can be integrated, i.e., interpreted together

These enable users to

- (a) Accurately find a target
- (b) Interpret each axis individually
- (c) Reach a target on the shortest path,

which are critical demands on a navigation system.

Even though these studies treat the topic well, formulate problems, necessities and suggest solutions, they lack experimental evaluation. Pioneering work has been done by [18]. Here, different strategies were applied to sonify the distance of a target using psychoacoustic quantities, such as loudness, pitch, brightness, beating and inharmonicity. After the participants carried out a training, the approaches were examined in an experiment with several tasks. A target point lay between 20 and 27.5 cm to the right of a starting point on a tablet computer. One task was to find the target as quickly as possible with a stylus. In another scenario, participants had to find the target point as precisely as possible. All participants

http://doi.org/10.21785/icad2018.007

were able to find the sonified target locations. Even though only the distance along one dimension is sonified, i.e., the approach is not multi-dimensional but 0.5-dimensional, the results serve as a benchmark for our own study in which we examine navigation in a 2-dimensional space, i.e. including distance and direction along two Cartesian dimensions. Details on the results of their study are resented in Sec. 5 and compared to the results of our own study.

The remainder of this paper is structured as follows: first, the psychoacoustic auditory display is briefly introduced. Then, we present important results of our pilot study, followed by the description of our current experiment. The results are presented and discussed against the results of [18]. The conclusion is followed by a brief outlook on further developmental steps.

3. PSYCHOACOUSTIC AUDITORY DISPLAY

In contrast to many auditory displays, our psychoacoustic auditory display does not map orthogonal spatial dimensions to independent audio parameters, like amplitude and fundamental frequency. Instead, it maps the dimensions to independent perceptual auditory qualities through signal processing. In principle, the sound informs the user where the target lies, relative to the current location by means of one auditory stream with different perceptual auditory qualities. This means the mapping is user-centric. We describe the mapping principle in colloquial terms to make it understandable for non-experts in psychoacoustics. This terminology is the same that we have used to describe the auditory display to the participants in our experiment.

The mapping principle is illustrated in Fig. 1 showing the same three exemplary cursor-target-constellations from two perspectives. The graphics indicate the sound features for cursor locations when the target lies in the center (a) and for target locations when the cursor lies in the center (b). We presented both versions to the participants of our experiment while explaining the mapping principle to them. Details on both the psychoacoustic background and the technical implementation are out of scope of this paper but can be found in [19, 20, 21]. Demonstration videos of the auditory display for some simple trajectories can be found on the first author's YouTube channel[1].

The target is a circular region represented by pink noise. In Fig. 1 targets are represented by red circles. The horizontal direction of the target relative to the current location is mapped to the direction of pitch: when pitch rises, the target lies to the right, when it falls, to the left. The distance is mapped to the speed: the faster the rise or fall, the further away the target lies within the horizontal x-dimension. At the center of the target, pitch is constant. In psychoacoustic terms not pitch per se is increasing or decreasing; in fact, chroma [22] is altered either clockwise or anticlockwise, while height is kept constant. This is achieved by a so-called *Shepard-tone* [22], which creates the auditory illusion of an infinitely rising or falling pitch even though it is a cyclic repetition of a sweep sequence. The speed of pitch rise or fall is actually the cycle speed of the sequence repetition. This Shepard tone consists of octaves only, so the sound does not exhibit roughness or beating and both brightness and loudness are equal for each pitch. The highest cycle speed is lower than 10 Hz, so that physical cycle duration equals perceived cycle duration [6, ch. 12].

The vertical y-dimension is divided in two. When the target lies below the current location the original Shepard tone is manipulated to sound rough: the further away, the rougher the sound. At the target height, the sound is smooth. When the target lies above the current location, the sound will remain smooth, but beating, i.e., regular loudness fluctuation is audible. The further away, the faster the beating. However, even the fastest beating is so slow that it will not be perceived as roughness. At the target height, the loudness is steady. Since a very slow beating as well as a very subtle degree of roughness are barely audible, the target height is indicated by an additional audible click, which represents the x-axis in a target-centered coordinate system. In technical terms, all frequency components of the original Shepard tone act as carrier frequencies in a frequency modulation synthesis [23] to create the roughness impression. Here, the modulation frequency is high enough to create no audible vibrator effect but sidebands near the carrier frequency. The further the target lies below, the higher the modulation depth, i.e. the higher the number and the amplitudes of the sidebands. Perceptually, this does not only increase the degree of roughness but also makes the sound increasingly inharmonic and noisy [24]. The modulation depth is restricted to the region in which the sound preserves high pitch salience [25], so that the pitch metaphor for the x-dimension can even be interpreted clearly at the highest degree of roughness. The beating impression is achieved by amplitude modulation with a low depth. The higher the target, the higher the amplitude modulation frequency. However, it is kept not only well below 15 Hz, where the impression of beating starts fading into the perception of roughness [6, ch. 11], but even below 10 Hz, where perceived duration equals physical duration [6, ch. 12]. When the central target location is reached, a smooth sound with steady pitch and loudness is heard, together with background noise.

The sonification and the additional elements, i.e., the click and the noise, are segregated auditory streams [26]. According to [26] it is easy to notice the presence of one auditory stream while listening to another auditory stream but almost impossible to recognize and follow two auditory streams at once. This is why the triggered click and noise only carry binary information: "central target height reached" and "target reached", respectively.

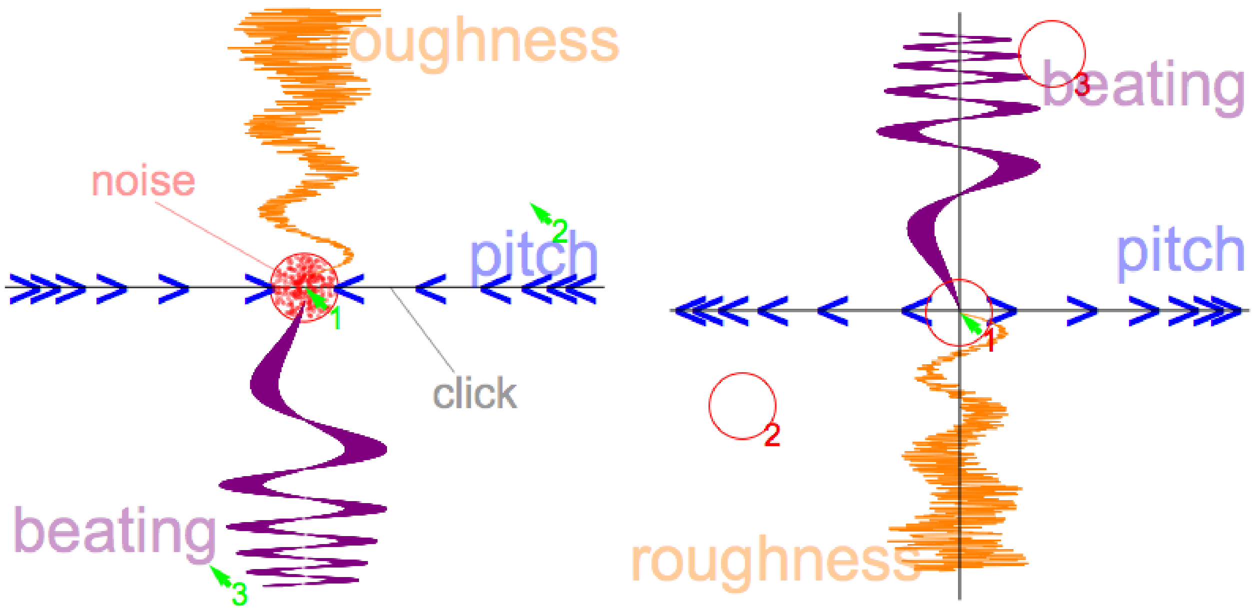
If the psychoacoustic principles are implemented correctly, the auditory display fulfills the three necessities mentioned above:

a) The sonification is interactively interpretable, which α) enables users to accurately find targets. The interpretability is mainly a matter of b) and c).

b) The sonification is perceived as one auditory stream, i.e., one sound with several characteristics, which are interpreted to derive the target location. At most locations and motions in the two-dimensional space, pitch alterations co-occur with alterations of beating frequency or roughness intensity. But these perceptual auditory qualities are largely independent from each other. Consequently, β) users should be able to either hear out information about the axes separately, and navigate axis-by-axis. For example, they could interpret the speed and direction of pitch change first, to derive the distance and direction along the x-axis. After navigating to the sonified x-coordinate, they interpret the degree of roughness or the speed of beating to reach the corresponding y-coordinate.

c) Alternatively, users could mentally combine the sound characteristics; e.g., a subtly rough sound with a quickly rising pitch means that the target is slightly below and far to the right. This way, they could γ) interpret the location of the target and approach it on the shortest path. The ability to interpret both dimensions together (shortest path) is what makes a multi-dimensional navigation superior to a succession of one-dimensional navigations

[1] `https://tinyurl.com/ycwmdh8r`.



(a) Target-Centered Perspective

(b) User-Centered Perspective

Figure 1: Mapping principle describing the sound for different locations of the user (cursors) with the target in the center (a) and of multiple targets (circles) with the user in the center (b).

(axis-by-axis).

4. METHOD

We carried out a pilot study to evaluate the learnability and interpretability of the pure sonification. Then, we examined the suitability of the whole auditory display for navigation in an interactive experiment.

4.1. Pilot Study

In a pilot study [19, 20, 27], we examined whether users are able to interpret the psychoacoustic sonification after a short explanation.

7 participants were introduced to the psychoacoustic sonification in a five-minute explanation with demo sounds. In the main experiment 19 sounds were played to them successively. These were 7 s-long sonifications of a static target, followed by 7 s of silence. Within these 14 s the participants had to assign the sound to one of 16 fields on a map, shown in Fig. 2. Even though one participant performed at chance level, an overall average of 41% of the sounds have been assigned to the correct field, 83% to the correct quartile. The figure indicates how often the individual field and each quartile have been assigned correctly. The outcome of this study provided evidence that people are able to quickly learn to interpret the sonification in a passive listening test. Based on this pilot study, some improvements of the sonification could be realized.

The results of this study motivated us to carry out an interactive experiment to evaluate the suitability of the whole auditory display for a navigation task. In the pilot study the participants were passive and had 7 s to listen and additional 7 s to think about the meaning of the sound. However, in an interactive navigation situation, the user is active and the sonification dynamic. This scenario is the basis of the navigation experiment.

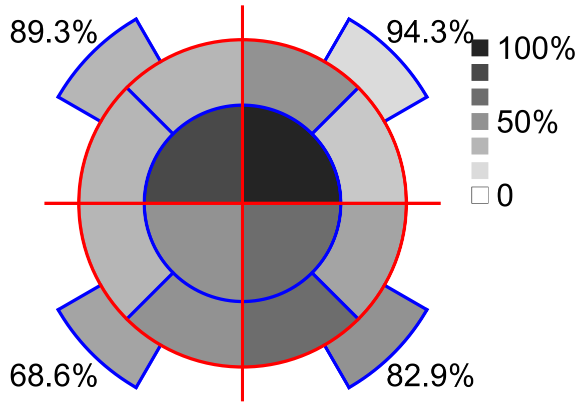


Figure 2: Map with 16 fields and percentages of correctly assigned field (gray level) and quartile (numbers in the corners).

4.2. Navigation Experiment

The psychoacoustic auditory display for navigation in two-dimensional space is evaluated in an experiment with novice users. Based on theoretic considerations and the results from the pilot study we hypothesize that a) the auditory display is interactively interpretable, b) the two sonified dimensions are orthogonal in perception c) the two dimensions are perceptually integratable, i.e., they are perceived as one auditory stream with several characteristics. If so, users were able to α) accurately find the targets, β)

analyze the sonification axis-by-axis, and γ) approach the target on the shortest path. So, we analyze the cursor trajectories of the participants to explicitly examine α), β), and γ) to implicitly provide evidence for a), b), and c), which we consider as a necessity for auditory navigation and as the main achievement of our psychoacoustic auditory display. A brief explanation of the experiment and initial findings can be found in [28]. A comprehensive paper about the experiment will be available soon in [21].

4.3. Setup

The experiment setup is illustrated in Fig. 3. Participants sat in front of a monitor and use a computer mouse to move a visible cursor on a screen towards an invisible target. The described auditory display guided them. The screen is located in the x-y-plane.

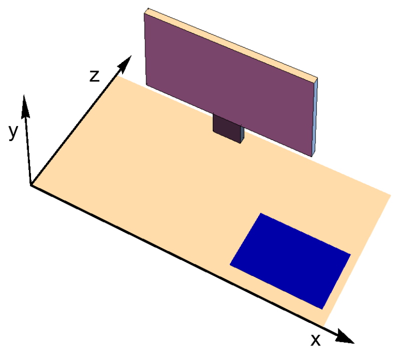


Figure 3: Experiment setup. A mouse is moved over a table to control a visible cursor on a otherwise blank screen.

4.4. Training

To get familiar with the auditory display, the experiment started with a training, always containing a verbal explanation, followed by an interactive demonstrator. The 7 demonstrators are depicted in Fig. 4. Since the participants were not familiar with psychoacoustic terms, we used more colloquial language. First, we played the target noise to the candidates. In demonstrator 1 they could move the cursor to trigger the target sound that appeared when a visible icon was reached. Next, we explained the pitch metaphor to them. In demonstrator 2 the candidates moved the cursor along the visible x-axis with the target in the center. They were advised to stop the mouse motion from time to time to listen closely to the sonification of a constant cursor-target-constellation and not only to the sonification of variable cursor-target-constellations, i.e., motion. Next, we explained the vertical dimension, i.e., the beating for targets above and the roughness for targets below the cursor, and the audible click at the target height. In demonstrator 3 candidates could move the cursor along the y-axis and interrupt their motion from time to time to get a feeling for a variable and a static cursor-target constellation. In demonstrator 4 candidates could move the mouse up and down along three additional vertical lines to the left and three to the right of the y-axis to experience the interactive beating/roughness dimension at different constant rates of pitch change. In demonstrator 5 they could move along 7 horizontal lines to experience the interactive pitch mapping at different constant beating rates or degrees of roughness. What followed was demonstrator 6 where candidates could activate single rows or columns out of the 7 horizontal and 7 vertical lines. In demonstrator 7 candidates could freely move through two-dimensional space with a fixed target in the center. Here, they were free to carry out diagonal or circular motions, zigzag lines or alike.

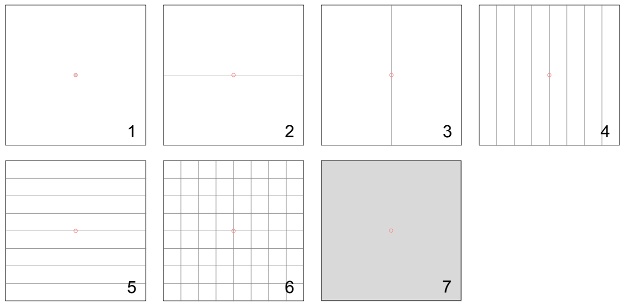


Figure 4: 7 demonstrator screens for the training. The circle represents the target. The sonification is active on/in the gray lines/area.

The training was concluded with an intermediate test, which resembles the test in the previous experiment. The candidates had to assign 10 sounds to the corresponding field on a map with 8 fields. The sonified targets lay in the center of the corresponding field. These fields were located slightly to the left/right and far up/down or far to the left/right and slightly up/down, illustrated in Fig. 5. Then, they described the sound characteristics: *Is pitch going up or down? Quickly or slowly?*, *Is the sound rough? Slightly or heavily? Or is it beating? Quickly or slowly?* After their description, candidates were allowed to revise their decision on the field assignment. Candidates passed the test when they allocated 5 out of 10 sounds correctly and at least 8 in the correct quartile. 18 out of 25 candidates passed the test and participated in the main experiment.

4.5. Task

For the main experiment, 16 targets with a radius of $b = 5$ mm were distributed equally on a square field with a side length of $a = 20$ cm on a monitor. The participants' task is to find 20 invisible targets as quickly as possible and click on them. First, the 16 targets appeared in random order, then, 4 random targets appeared again. The cursor started in the center of the screen and was moved by the participant to the anticipated target location. Here, the participant performed a click, which saved the mouse trajectory to a file, reset the cursor to the center of the screen, and loaded the next target. Participants were not informed whether they actually hit the target.

5. RESULTS AND DISCUSSION

Two quantitative measures serve to test the hypothesis that participants were able to find the target accurately due to the psychoacoustic auditory display. First, hits were counted, i.e., how frequently users found the targets. Second, the trajectories were analyzed in terms of the time to reach the target. Shorter duration for nearer targets indicates that the targets were found by interpreting

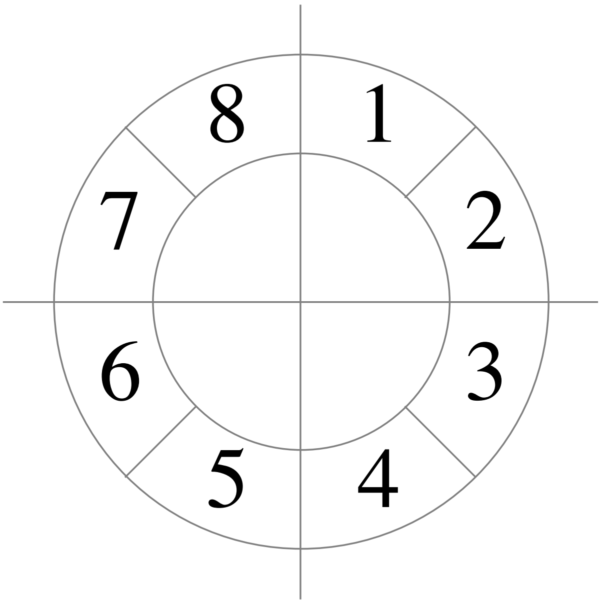


Figure 5: Map showing 8 fields in 4 quartiles.

the auditory display and not by systematically or randomly scanning the whole two-dimensional space. Qualitative, visual inspection of the trajectories provides additional evidence for or against a systematic or random scan of the whole space. The visual inspection also serves to test hypotheses b) and c). If participants navigate towards the target axis-by-axis, this is evidence that they are able to interpret each auditory dimension separately. If trajectories approximate the shortest path, this is evidence that participants are able to combine the two auditory dimensions.

Statistics of individual targets are summarized in Figs. 6 and 7. The large disks are centered at their corresponding target. The cross represents the spatial axes. In Fig. 6 the numbers indicate the mean hit rates $\pm$ the estimated standard errors. For a better overview disk shading redundantly indicates how frequently a target was hit. The darker the disk, the higher the hit-rate. The numbers in the corners give the mean hit-rate for the respective quartile. Most targets were hit in 90 to 100% of all trials with an average hit rate of 91.8%. One-way ANOVA revealed no significant effect of quartile on the hit-rate ($F(3, 12) = 0.85$, $p = 0.49$). This implies that the combinations of chroma change with beating and roughness were interpretable similarly well.

In Fig. 7 the number and the disk shading indicate the mean time to reach the target. The darker the disk, the longer the needed time. The numbers in the corners give the mean duration for the respective quartile. It took participants about 20.9 s to reach a target, ranging from 14.7 to 31.2 s. We found no significant relationship between the hit rate and the time to reach the target. A general trend can be observed that the further away the target, the longer it took participants to reach it. However, even though the outermost targets are about 3.5 times further away than the nearest, it took participants only twice as long to reach them. So in relation to the distance, near targets were reached relatively slower. On average, targets in all quartiles were reached within a similar amount of time. One-way ANOVA revealed no significant effect of quartile on the time to reach the target ($F(3, 12) = 1.23$, $p = 0.34$). This supports the finding from the hit-rate, i.e., that the combinations of chroma change with beating and roughness were interpretable similarly well.

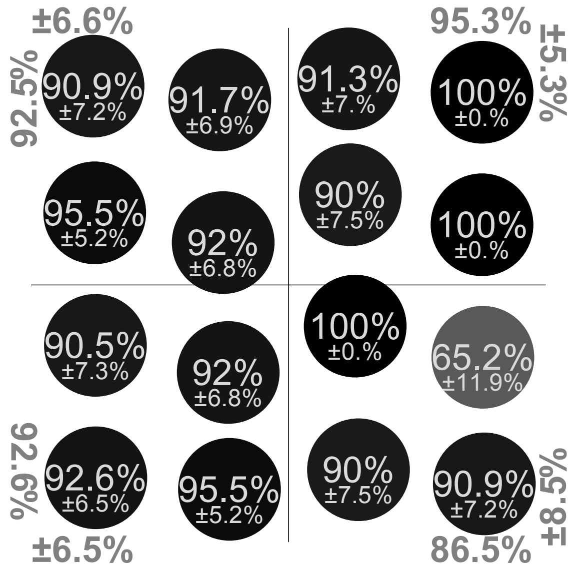


Figure 6: Hit rate $\pm$ estimated standard error for each target and the four quartiles.

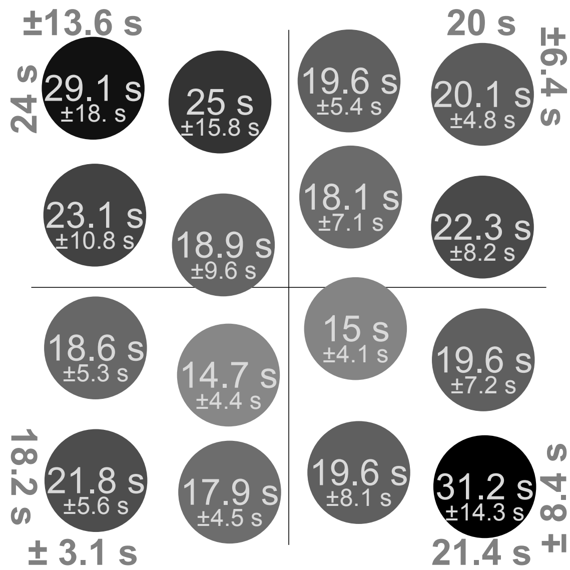


Figure 7: Mean time to reach each target $\pm$ standard deviation for each target and the four quartiles.

Statistics of the 18 individual performances are summarized in Figs. 9 and 8. The participants hit between 75% and 100% of the targets. On average, it took participants between 8 and 34 s to reach the targets, except for one individual, who needed no less than 54 s on average. We found no significant linear relationship between the hit-rate and the needed time. We observe that the longer the mean time to reach the target, the larger the standard deviation. The two correlate significantly (Pearson's correlation coefficient $r = 0.934347$, $F = 109.99$, $p = 1.4108 \times 10^{-8}$). This rather common relationship states that the slowest participants are not consistently slow. Their performance varies stronger, which may indicate that they are more insecure. After the experiment, some participants confessed that they confused left with right and up with down from time to time. One reason for that may be our misleading graphical representations of the auditory display during the training. We presented Fig. 1 (a) to them, which describes the sound when the target is in the center and the cursor is somewhere else in 2-dimensional space. When the cursor is far to the right and above the target, the sound exhibits fast pitch decrease and roughness. But we also presented Fig. 1 (b) to them, which describes the sonification when the cursor is in the center and the target is somewhere else in 2-dimensional space. When the target is far to the right and above the cursor, the sound will exhibit fast pitch increase and beating. These seemingly contradictory representations may have been the source of confusion. We hypothesize that an unambiguous explanation of the mapping principle and longer familiarization with the sonification metaphors and more experience with the auditory display may improve and stabilize the users' performance and their resoluteness.

The observed training effect supports this hypothesis. We compared the mean value of the first 6 with the last 6 valid runs for all users. Paired samples t-test revealed that the time to reach the target significantly improved by 9.8 s from 26.3 ± 8.3 s to 16.5 ± 3.0 s ($t(17) = 2.964947$, $p = 0.0043395$).

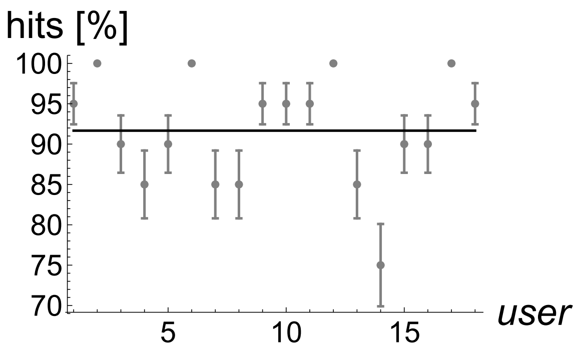


Figure 8: Hit-rate (mean value and estimated standard error) for each participant (user).

Some exemplary trajectories are plotted in Fig. 10. Here, the circles indicate the exact extent of the target in relation to the two-dimensional space. At least one trajectory is plotted for each of the 16 targets. It can be seen that some paths approach the target axis-by axis, either finishing one axis after the other, or interchanging between them. Other trajectories aim relatively straight towards the target. It can be observed frequently that participants oscillate around the target height to repeatedly trigger the click to confirm that they are still at the correct height. Even comparably long trajectories are still targeted and do not resemble a systematic or random scan of the whole two-dimensional field. The trajectories of the worst performances, i.e., the six longest times to reach a target are plotted in Fig. 11. Here, some axis-by-axis motions and oscillations around the target height can be observed. These trajectories do not seem to scan the whole field. They only cover a comparably small part of the field and rather tend to concentrate on rough target x and/or target y-coordinate.

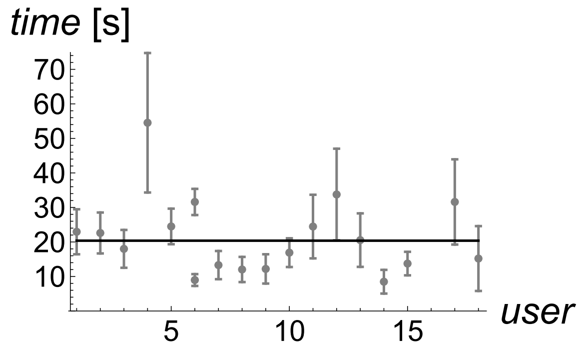


Figure 9: Performance (mean value and standard deviation) of each participant (user).

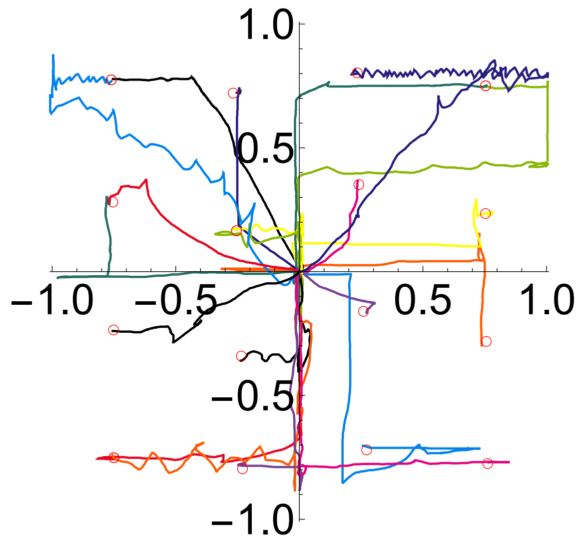


Figure 10: Typical trajectories. Some participants solve the x-axis first, others the y-axis. Some switch between motions along the two. Some trajectories approach the target relatively straight. Many trajectories oscillate around the target height to repeatedly trigger the click as confirmation that their height is still correct.

The data confirms our hypotheses that the auditory display enables users to reliably find invisible targets, analyze the sound axis by axis or integrate the information about both axes to aim for the shortest path. The participants find the targets faster and more well-aimed than with a random scan of the whole space. This is true for targets in all quartiles, i.e., the combination of pitch with beating and the combination of pitch with roughness work equally well.

Many auditory displays lack experimental evaluation [15]. Therefore, it is difficult to quantify the benefit of our psychoacoustic auditory display, because no benchmark exist. The only benchmark is the study by [18] for a half-dimensional guidance task. Note that one should be wary of the comparison results, due to different experiment setups, trainings, scenarios, tasks and populations: The participants of our study took 20.9 s to find a target that covered $2 \times \pi b/a^2 = 0.031\%$ of a two-dimensional space. When their task was to find the target as precisely as possible, users

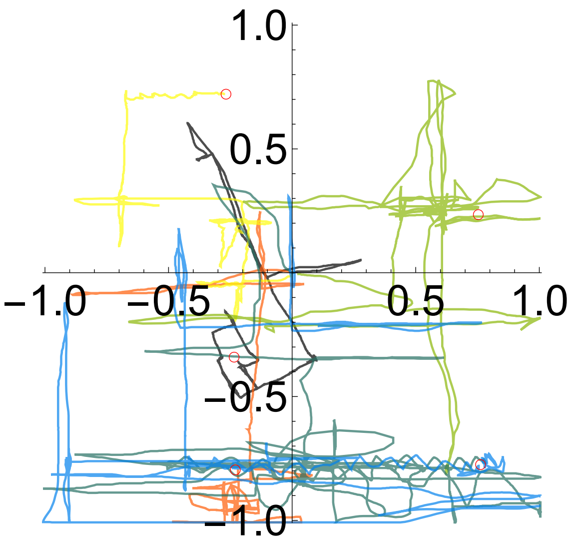


Figure 11: Trajectories of the 6 longest times to reach a target. They occupy only a small region of the field.

in [18] took a similar amount of time to find a target with a precision of 0.97% $\pm 0.43\%$ along a half-dimensional space, i.e., along one line to the right. Their best methods achieved a precision of 0.1% and 0.2%. When their task was to find the target as quickly as possible the participants took 4.5 ± 0.1 s to approach the target with an error of 10.2 ± 1.9 mm, i.e., $3.68\% \pm 0.68\%$. Obviously, our psychoacoustic auditory display enabled users to find a smaller target in less time even though a two-dimensional space is sonified instead of a half-dimensional space. However, participants in our study took roughly 30 minutes of training, including explanation, demonstrators, and intermediate test. Participants in [18] only took 7 ± 2 trials, on average, to train the task.

6. CONCLUSION

In this paper a psychoacoustic auditory display for navigation in two-dimensional space has been presented and experimentally evaluated. The sonification principle is a mapping of relative direction and distance to a combination of direction and speed of pitch change with the speed of beating or the degree of roughness. The approach considers psychoacoustics in terms of chroma [22], pitch and pitch salience [25] [6, ch. 5], perceived duration [6, ch. 12], beating and roughness [6, ch. 11] [24], inharmonicity and noisiness [24].

After only 30 minutes of explanation and training, targets were a) accurately found way faster and more targeted than a systematic scan of the whole two-dimensional space. Users were able to b) approach the target axis-by-axis or c) on the shortest path. These observations as evidence that our psychoacoustic auditory display is α) interactively interpretable, that β) the axes are orthogonal in perception, and γ) the sonification is perceptually integrated as one auditory stream and thus both axes can be interpreted together to derive the target direction and distance. The performance of our participants is comparable or even better than reported for a half-dimensional navigation task in [18]. Furthermore, a highly significant training effect could be observed, indicating that the users improve their performance with further practice, so we expect even better results for more experienced users.

7. OUTLOOK

Naturally, many navigation tasks take place in 3-dimensional spaces. We included constants in our signal processing, which could act as variables to map the third dimension to perceived *brightness* [29] and *fullness*, *volume* or *sonority* [29], [30], [31, p. 31]. Interactive experiments could serve to validate that this dimension is well interpretable and orthogonal to the other two.

The exact mapping from the physical input to the audio parameters which create the desired perceptual output is based on general knowledge from the field of psychoacoustics. The mapping from the physical distance to the speed of pitch change and of beating is linear, as it is kept within the range in which the perception of duration is reported to be linear [6, ch. 12]. The roughness-mapping is a combination of a linear and an exponential term, which seemed necessary in consideration of roughness perception as reported in the literature [24, 32, 6] to enable users to distinguish several magnitudes in all regions from a low to a large distance. Here, implementing psychoacoustic roughness models as in [24, 32] to derive the optimal mapping function has the potential to perfect the mapping towards perceived linearity and continuity.

To date, psychoacoustic models tend to be valid for static sounds, like [24, 6, 25]. Others have been validated with dynamic sounds, like notes of musical instruments [32, 29]. However, there is an urgent need for psychoacoustic models in interactive scenarios with dynamic sounds and active users [11, 12], not only in the field of auditory display research but also in musicology, audiology and psychoacoustics. The psychoacoustic auditory display could serve as a tool to develop such psychoacoustic models in interactive scenarios.

Our experiment served as a proof-of concept and highlights its strengths and weaknesses. A direct comparison of auditory displays for navigation may be an interesting topic for the future. However, a drawback of a direct comparison of methods is that a number of sonification principles have to be learned by users. This is very time consuming and may cause fatigue and confusion of mapping principles. Therefore, we analyzed the users' trajectories by several means, which could serve as benchmarks. These benchmarks enable to compare navigation performance between different studies. Details can be found in [21].

8. ACKNOWLEDGMENT

The authors thank David Black, Frederik Nagel, Thomas Sporer, Robert Mores and Stefan Weigand for fruitful discussions about the psychoacoustic principles, auditory display evaluations, and future research. We also thank Wasutorn Sanamchai, who programmed the software used in our experiment. We are grateful to the two anonymous reviewers whose ideas and formulations improved our paper.

9. REFERENCES

[1] A. Lundkvist, R. Johnsson, A. Nyknen, and J. Stridfelt, "3d auditory displays for parking assistance systems," *SAE Int. J. Passeng. Cars Electron. Electr. Syst.*, vol. 10, pp. 17–23, 04 2017. [Online]. Available: https://doi.org/10.4271/2017-01-9627

[2] D. S. Brungart and B. D. Simpson, "Design, validation, and in-flight evaluation of an auditory attitude indicator based on pilot-selected music," in *Proc. 14th International Conference on Auditory Display (ICAD2008)*, Paris, Jun 2008, p. 8 pages. [Online]. Available: http://hdl.handle.net/1853/49897

[3] A. Vasilijevic, Z. Vukic, and N. Miskovic, "Teleoperated trajectory tracking of remotely operated vehicles using spatial auditory interface," in *IFAC PapersOnLine*, vol. 49, no. 23, 2016, pp. 079–102.

[4] D. Black, J. Hettig, M. Luz, C. Hansen, R. Kikinis, and H. Hahn, "Auditory feedback to support image-guided medical needle placement," *International Journal of Computer Assisted Radiology and Surgery*, pp. 1–9, 2017. [Online]. Available: http://dx.doi.org/10.1007/s11548-017-1537-1

[5] J. E. Anderson and P. Sanderson, "Sonification design for complex work domains: Dimensions and distractors," *Journal of Experimental Psychology: Applied*, vol. 15, no. 3, pp. 183–198, Mar 2009. [Online]. Available: http://dx.doi.org/10.1037/a0016329

[6] H. Fastl and E. Zwicker, *Psychoacoustics. Facts and Models*, 3rd ed. Berlin, Heidelberg: Springer, 1999. [Online]. Available: https://doi.org/10.1007/978-3-540-68888-4

[7] G. Kramer, "Some organizing principles for representing data with sound," in *Auditory Display: Sonification, Audification, and Auditory Interfaces*, ser. Santa Fe Studies in the Science of Complexity, Proc. Vol. XVIII, G. Kramer, Ed. Reading, MA: Addison-Wesley, 1994, pp. 223–252.

[8] S. Barrass, "A perceptual framework for the auditory display of scientific data," in *Proc. 2nd International Conference on Auditory Display (ICAD1994)*, Santa Fe, Nov 1994, pp. 131–145. [Online]. Available: http://hdl.handle.net/1853/50821

[9] S. Smith, "Representing data with sound," in *Proc. IEEE Visualization*, Piscataway (NJ), 1990.

[10] S. M. Williams, "Perceptual principles in sound grouping," in *Auditory Display: Sonification, Audification and Auditory Interfaces*, G. Kramer, Ed. Reading (MA): Addison Wesley, 1994, pp. 95–126.

[11] B. N. Walker and G. Kramer, "Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making," in *Ecological Psychoacoustics*, J. G. Neuhoff, Ed. Amsterdam: Elsevier, 2004, ch. 6, pp. 149–174.

[12] S. Ferguson, D. Cabrera, K. Beilharz, and H.-J. Song, "Using psychoacoustical models for information sonification," in *Proc. 12th International Conference on Auditory Display (ICAD 2006)*, London, Jun 2006. [Online]. Available: http://hdl.handle.net/1853/50694

[13] J. P. Bliss and R. D. Spain, "Sonification and reliability — implications for signal design," in *Proc. 13th International Conference on Auditory Display (ICAD 2007)*, Montral, Jun 2007, pp. 154–159. [Online]. Available: http://hdl.handle.net/1853/50028

[14] A. Hunt and T. Hermann, "Interactive sonification," in *The Sonification Handbook*, J. G. N. Thomas Hermann, Andy Hunt, Ed. Berlin: COST and Logos, 2011, ch. 11, pp. 273–298. [Online]. Available: http://sonification.de/handbook/

[15] N. Degara, F. Nagel, and T. Hermann, "Sonex: An evaluation exchange framework for reproducible sonification," in *Proc. 19th International Conference on Auditory Display (ICAD 2013)*, Lodz, Jul 2013. [Online]. Available: http://hdl.handle.net/1853/51662

[16] T. Bovermann, J. Rohrhuber, and A. de Campo, "Laboratory methods for experimental sonification," in *The Sonification Handbook*, T. Hermann, A. Hunter, and J. G. Neuhoff, Eds. Berlin: COST and Logos, 2011, ch. 10, pp. 237–272. [Online]. Available: http://sonification.de/handbook/

[17] D. Arfib, J. Couturier, Kessous, L., and V. Verfaille, "Strategies of mapping between gesture data and synthesis model parameters using perceptual spaces," *Journal of Organised Sound*, vol. 7, no. 2, pp. 127–144, 2002.

[18] G. Parseihian, C. Gondre, M. Aramaki, S. Ystad, and R. Kronland-Martinet, "Comparison and evaluation of sonification strategies for guidance tasks," *IEEE Trans. Multimedia*, vol. 18, no. 4, pp. 674–686, April 2016.

[19] T. Ziemer, D. Black, and H. Schultheis, "Psychoacoustic sonification design for navigation in surgical interventions," *Proceedings of Meetings on Acoustics*, vol. 30, 2017. [Online]. Available: https://doi.org/10.1121/2.0000557

[20] T. Ziemer and D. Black, "Psychoacoustically motivated sonification for surgeons," in *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. (Suppl 1):1, Barcelona, Jun 2017, pp. 265–266. [Online]. Available: https://link.springer.com/article/10.1007/s11548-017-1588-3

[21] T. Ziemer, H. Schultheis, D. Black, and R. Kikinis, "Psychoacoustical interactive sonification for short range navigation," *(accepted for) Acta Acust. united Ac.*, vol. 104, 2018. [Online]. Available: http://www.ingentaconnect.com/content/dav/aaua

[22] R. N. Shepard, "Circularity in judgments of relative pitch," *J. Acoust. Soc. Am.*, vol. 36, no. 12, pp. 2346–2353, 1964. [Online]. Available: https://doi.org/10.1121/1.1919362

[23] J. Chowning and D. Bristow, *FM Theory & Applications. By Musicians for Musicians*. Tokyo: Yamaha Music Foundation, 1986.

[24] W. Aures, "Berechnungsverfahren für den sensorischen wohlklang beliebiger schallsignale (a model for calculating the sensory euphony of various sounds)," *Acustica*, vol. 59, no. 2, pp. 130–141, 1985.

[25] E. Terhardt, G. Stoll, and M. Seewann, "Algorithm for extraction of pitch and pitch salience from complex tonal signals," *J. Acoust. Soc. Am.*, vol. 71, no. 3, 1982. [Online]. Available: http://doi.org/10.1121/1.387544

[26] A. S. Bregman, *Auditory Scene Analysis*. Massachusetts: MIT Press, 1990.

[27] T. Ziemer and D. Black, "Psychoacoustic sonification for tracked medical instrument guidance," *J. Acoust. Soc. Am.*, vol. 141, no. 5, p. 3694, 2017. [Online]. Available: http://dx.doi.org/10.1121/1.4988051

[28] T. Ziemer and H. Schultheis, "Perceptual auditory display for two-dimensional short-range navigation," in *accepted for: Fortschritte der Akustik (DAGA2018): 44.*

Deutsche Jahrestagung für Akustik, Munich, Mar 2018. [Online]. Available: https://www.dega-akustik.de/publikationen/online-proceedings/

[29] J. W. Beauchamp, "Synthesis by spectral amplitude and "brightness" matching of analyzed musical instrument tones," *J. Audio Eng. Soc.*, vol. 30, no. 6, pp. 396–406, 1982.

[30] A. Schneider, "Perception of timbre and sound color," in *Springer Handbook of Systematic Musicology*, R. Bader, Ed. Berlin, Heidelberg: Springer, 2018, ch. 32, pp. 687–726. [Online]. Available: http://doi.org/10.1007/978-3-662-55004-5_32

[31] J. Meyer, *Acoustics and the Performance of Music*. New York, NY: Springer, 2009. [Online]. Available: http://doi.org/10.1007/978-0-387-09517-2

[32] M. Leman, "Visualization and calculation of the roughness of acoustical musical signals using the synchronization index model (sim)," in *Proc. COST G-6 Conference on Digital Audio Effects (DAFX-00)*, Verona, Dec 2000.